




cambridge.org/bil

A. I. Pérez^{1,2} , E. Schmidt³ and I. M. Tsimpli²

¹Department of Experimental Psychology, University of Granada, Granada (Spain); ²Theoretical and Applied Linguistics, University of Cambridge, Cambridge (UK) and ³Cambridge Assessment English, University of Cambridge, Cambridge (UK)

Research Article

Cite this article: Pérez, A.I., Schmidt, E., & Tsimpli, I.M. (2023). Inferential evaluation and revision in L1 and L2 text comprehension: An eye movement study. *Bilingualism: Language and Cognition*, 1–14. <https://doi.org/10.1017/S136672892300069X>

Received: 13 May 2022
Revised: 11 September 2023
Accepted: 12 September 2023

Keywords:
L2 text comprehension; inferential evaluation; revision; inhibitory control; linguistic proficiency

Corresponding author:
Ana I. Pérez Muñoz
E-mail: anaipe@ugr.es

Abstract

Text comprehension frequently demands the resolution of no longer plausible interpretations to build an accurate situation model, an ability that might be especially challenging during second language comprehension. Twenty-two native English speakers (L1) and twenty-two highly proficient non-native English speakers (L2) were presented with short narratives in English. Each text required the evaluation and revision of an initial prediction. Eye movements in the text and a comprehension sentence indicated less efficient performance in the L2 than in L1 comprehension, in both inferential evaluation and revision. Interestingly, these effects were determined by individual differences in inhibitory control and linguistic proficiency. Higher inhibitory control reduced the time rereading previous parts of the text (better evaluation) as well as revisiting the text before answering the sentence (better revision) in L2 comprehenders, whereas higher proficiency reduced the time in the sentence when the story was coherent, suggesting better general comprehension in both languages.

1. Introduction

Different from lexical or sentence reading, successful text comprehension requires the construction of a coherent and accurate mental representation by which information from the text is integrated with reader's prior knowledge, building a situation model (Kintsch & van Dijk, 1978). Several high-level comprehension processes underlie the ability to build the situation model. For instance, as the text unfolds, previous story information and (more frequently) plausible interpretations generated by the reader (i.e., inferences) can become no longer relevant or plausible after encountering new pieces of information, and thus require updating. This might be especially challenging during second language (L2) comprehension. The present research expands a prior study investigating high-level cognitive processes during native (L1) text comprehension (Pérez et al., 2016), to L2 processing, and whether they are predicted by individual differences in cognitive control and/or linguistic proficiency. We will now discuss the processes of interest in relation to L1 and L2 comprehension.

1.1. Inference making

An essential comprehension process is inference making, which is the ability to extract and/or connect ideas that have not been explicitly referred to, facilitating coherence and text integration (e.g., Cain & Oakhill, 1999; Cook & O'Brien, 2017). Prediction is a relevant type of inference for text comprehension (for a broader discussion on different types of inferences see Pérez et al., 2014), as it helps to anticipate incoming story information by pre-activating the linguistic representation of a concept after combining text content and readers' prior knowledge (Pickering & Gambi, 2018). For example, predictable words like "shark" in (1) are fixated for less time than the same word in less predictable contexts.

1) "He saw the black fin slice through the water and the image of sharks' teeth came quickly to his mind. He turned quickly toward the shore and swam for his life. The coast guard had warned that someone had seen a shark off the north shore of the island. As usual, not everyone listened to the warning." (Ehrlich & Rayner, 1981)

Interestingly, the topic of predictive processing has been gaining importance in the bilingual literature. In a recent review, Schlenter (2022) proposes that quantitative differences (i.e., later onsets and/or weaker effects in the L2 compared to the L1) are more frequent than qualitative differences (i.e., effects in the L2 are absent or different from the L1) in predictive processing. Specifically, qualitative differences in prediction are usually found under morphological or phonological cues' manipulations (see e.g., Ito et al., 2018; Martin et al.,

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

2013), whereas quantitative differences are mainly related to semantic linguistic aspects such as semantically biased verbs (Chun & Kaan, 2019), semantic numeral classifiers (Mitsugi, 2020), polarity adverbs (Mitsugi, 2022), verbs selecting an animate object (Schlenter & Felser, 2021), words in highly constrained sentences (Dijkgraaf et al., 2019), as well as the use of implicit causality at the discourse-level (Kim & Grüter, 2021). Altogether, this literature suggests that predictive processing may be slower and/or weaker in the L2 compared to the L1, especially when conceptual (semantic) information is required. In addition, this is consistent with a theoretical hypothesis claiming that, compared to native speakers, non-native speakers have a “Reduced Ability to Generate Expectations” (RAGE hypothesis; Grüter et al., 2014, 2017; see Dijkgraaf et al., 2017, for contradictory evidence). A more recent reformulation of the RAGE hypothesis (Grüter & Rohde, 2021) specifies that language differences could come not just from a reduced ability of L2 speakers to engage in prediction (deficient processing view), but also from a reduced utility of prediction when L2 speakers perceive that its generation will cause more costs than benefits (adaptive processing view).

1.2. Evaluation

A second important high-level comprehension process is evaluation, which refers to our ability to be aware of possible conflicting information, inconsistencies, or any unexpected information with respect to previous parts of the text. Evaluation is typically assessed by means of an inconsistency-detection paradigm, where a sentence or text such as (1) includes either an expected (“shark”) or unexpected (e.g., “orca”) but still plausible concept. Skilled comprehenders increase their processing time (e.g., longer reading times, longer fixations durations) in the unexpected compared to the expected condition (e.g., Cain et al., 2004; Joseph et al., 2021), indicating good comprehension evaluation. Moreover, evaluation can also occur at the inferential level. For instance, Kaakinen et al. (2014) observed that adult readers made more regressions (looks back) from a critical sentence into the previous context when ironic compared to non-ironic information had been presented¹, signalling their ability to evaluate well unexpected inferential information at the situation model level during online native comprehension. Nonetheless, inferential evaluation can be demanding especially if it occurs in the second language (e.g., Kaan & Grüter, 2021).

Similar to native comprehenders, young adults seem to evaluate local coherence by detecting inconsistencies between adjacent sentences in the L2. However, they frequently fail to do so when inconsistencies are presented across distant sentences, thus failing to evaluate global coherence during L2 comprehension (e.g., Morishima, 2013; Ushiro et al., 2016, 2021). In an eye-tracking study, Ushiro et al. (2016) assessed Japanese–English bilinguals’ eye movements while they were reading English texts containing inconsistencies about the protagonists that disrupted coherence either locally or globally. Similar to Kaakinen et al. (2014), they observed more regressions from the inconsistencies to previous regions of the story only in the local (but not global) condition in L2 comprehension. Moreover, Pérez et al. (2019) assessed inferential evaluation at the situation model level in bilingual young adults reading stories in both L1-Spanish and L2-English. They observed longer reading times in the unexpected (update) condition compared to the expected or neutral (non-update and neutral) conditions in both L1 and L2 – however,

this effect was reduced in the L2, suggesting that the evaluation process was less efficient in the non-native language. To sum up, these effects indicate that L2 readers may experience difficulties evaluating comprehension at the situation model level compared to L1 comprehenders.

1.3. Revision

Once unexpected information is noticed, readers are required to engage in repair processes such as updating information to resolve the conflict. Revision requires not only the activation of alternative information, but also the rejection of the no longer plausible interpretation (Rapp & Kendeou, 2007). Using our previous example (1), native comprehenders usually update their situation model by activating the new concept “orca” once it is presented (e.g., de Vega, 1995; Rapp & Taylor, 2004). However, the initial prediction (“shark”) is often maintained, causing interference and disrupting comprehension (e.g., Blanc et al., 2011; Rapp & Kendeou, 2009). Accordingly, the Knowledge Revision Comprehension framework (KReC; Kendeou & O’Brien, 2014) proposes five processes that are required during the revision of a situation model: 1) ENCODING, in which the initial information (in our case a prediction) is encoded in long-term memory; 2) PASSIVE ACTIVATION, by which that information has the potential to become activated by any related content maintained in working memory; 3) CO-ACTIVATION, if an alternative information is encoded, it causes the activation of both the initial and this new information at the same time; 4) INTEGRATION, the new information makes contact with the initial information triggering revision of the mental representation; and 5) COMPETING ACTIVATION, which helps to draw activation away from the (now) no longer appropriate initial information, reducing interference.

In the eye movement study conducted by Pérez et al. (2016), young native English-speaking adults were presented with short narratives as in (2) containing either an expected (“oven”) or an unexpected (“grill”) word. Longer go-past time (the sum of all fixations from first entering a region from the left to exiting it from the right, including later re-reading previous parts of the text), in the unexpected compared to the expected target word, demonstrated efficient online inferential evaluation during L1 comprehension. More importantly, a comprehension sentence such as in (3) was presented below the text, simultaneously with the text. This sentence brought either a concept that was congruent with the expected word but incongruent with the unexpected word (“roasted”), or inversely, a concept that was congruent with the unexpected word but incongruent with the expected word (“barbecued”). Here, comprehenders took longer total time (the total duration of all fixations in a region, including first and second-past times) to read the congruent concept when the unexpected word had previously been presented in the story (“grill” → “barbecued”) compared to when the expected word (“oven” → “roasted”) had been presented. This processing increase was interpreted as the need to discard the no longer plausible initial inference to be able to confirm the alternative interpretation, which certainly reflects revision during native comprehension.

2) *It was the 25th of December and Sophie was back home. As a special treat, her father was making her a traditional Christmas dinner. The turkey was cooking, and it needed another hour in the oven/grill before it was done.*

3) *The turkey needed to be roasted/barbecued for one more hour.*

Pérez et al.'s (2016) study is crucial for the present research, as we used the same experimental paradigm to assess both L1 and L2 comprehension. Unfortunately, the literature on revision during L2 online text comprehension is not as extensive as for evaluation. However, Schleicher and Schwartz (2022) assessed discourse revision in highly proficient bilinguals comprehending expository texts in both L1-Spanish and L2-English. In that study, contradictory information was presented across two texts. The authors observed lower accuracy in questions which required revision when the to-be-updated information was presented in the L2, especially if the to-be-discarded information was read in the L1. This was interpreted as a result of interference caused by the reactivation of the initial information. This is in line with Pérez et al.'s (2019) results who also investigated inferential revision in highly proficient bilinguals, during L1-Spanish and L2-English text comprehension. Specifically, after presenting the relevant condition (i.e., neutral, non-update or update, see above), participants encountered a final sentence with a critical word ("*grill*") that was always inconsistent with the initial prediction ("*oven*"), but consistent with the alternative suggested in the unexpected (update) condition ("*barbecued*"). Bilinguals manifested larger N400 (i.e., a measure of the ability to integrate the new prediction into the situation model) coming from the expected and neutral (non-update and neutral) conditions compared to the unexpected (update) condition, when comprehending in both the L1 and the L2. Nevertheless, the N400 effect was smaller in the latter, indicating that L2 revision was qualitatively similar but quantitatively different compared to the L1. In sum, these studies suggest that revision is a very cognitively demanding process, especially because of the need to deal with interference caused by an initial interpretation. It affects both L1 and L2 comprehension, but once more has a stronger impact on the non-native language.

1.4. Cognitive control and linguistic proficiency

Importantly, some studies have shown that evaluation and (more consistently) revision are sometimes predicted by individual differences in several executive measures such as working memory or cognitive control, as well as by L2 linguistic proficiency in bilinguals' literature.

Only very few studies have demonstrated a relationship between executive function and the ability to detect inconsistencies in the L1 in young adults (e.g., St. George et al., 1997), suggesting L1 evaluation does not necessarily require executive control. In fact, evaluation has been considered a routine, passive, and nonstrategic process (Kendeou, 2014). This is also consistent with findings demonstrating that the detection of information that mismatches prior-knowledge stored in long-term memory is not associated with working memory and/or inhibitory control (Pérez et al., 2015, 2016, 2020). On the other hand, the literature on evaluation in a non-native language has yielded mixed results. Pérez et al. (2019) did not find individual differences in either cognitive control (AX-CPT task) or linguistic proficiency (composite score of L2 vocabulary, L2/L1 verbal fluency and L2 self-assessment measures) to be associated with bilinguals' evaluation performance (in neither L1 nor L2). In contrast, Zirnstein et al. (2018) assessed highly proficient Chinese-English bilinguals reading in their L2 (Exp. 2) by means of highly constrained sentences that led to a clear lexical prediction, and found that their ability to detect a mismatch in an unexpected word positively correlated with both cognitive control (AX-CPT task) and linguistic proficiency (L1 verbal fluency). Finally, a recent reanalysis of Pérez

et al.'s (2019) data has also demonstrated that working memory (L2/L1 operational span task) predicts inferential evaluation in both L1 and L2 comprehension (Pérez & Bajo, 2022). Therefore, even though evaluation does not seem to involve executive control in monolingual speakers (Pérez et al., 2015, 2016), this relationship is less clear in bilinguals, as some studies have shown no relationship with cognitive control (Pérez et al., 2019), while others have pointed at individual differences in inhibitory control (Zirnstein et al., 2018) and working memory (Pérez & Bajo, 2022), either in L1 or in L2 comprehension.

In contrast, evidence relating executive function to the process of revision during L1 comprehension is rather substantial (Pérez et al., 2015, 2016, 2020). For example, Pérez et al. (2016) found that lower verbal (but not visuospatial) working memory English speakers spent more (go-past) time revisiting the text (2) after reading the comprehension sentence (3), but only when the target word of the text region was unexpected ("*grill*"). This effect was interpreted as a less efficient ability of lower (compared to higher) span readers to draw activation away from the initial interpretation ("*oven*"), and therefore to successfully perform revision at the situation model. In a subsequent study, Pérez et al. (2020) included both working memory (backward digit recall task) and inhibitory control (flanker task) in the statistical model, and only the latter was associated with revision, suggesting this process is mainly based on an inhibitory mechanism. Finally, to the best of our knowledge, the only study investigating individual differences in L2 revision is Pérez et al. (2019). Here, they found that L2 revision (N400) was related to both cognitive control and linguistic proficiency (see measures above), suggesting that higher proactive control as well as higher L2 proficiency helped to reduce interference from the no longer plausible initial prediction. However, notice this contrast with the reanalysis performed by Pérez and Bajo (2022), where working memory (see above) predicted individual differences in revision only during L1 (but not L2) comprehension. All these findings indicate the importance of executive control and (more specifically) of inhibitory control in L1 revision, as well as the need to provide more evidence regarding individual differences in both cognitive control and linguistic proficiency in L2 revision.

1.5. The present study

Previous studies have been mainly conducted using sentences instead of texts, and they have targeted lower linguistic levels such as grammatical or phonological aspects, instead of doing so at the situation model level. Importantly, texts are processed incrementally (word by word, and sentence by sentence). As each new piece of information is encountered, readers must update their mental representation of the text, which makes the construction of the situation model a complex dynamic process (e.g., McNamara & Magliano, 2009). Moreover, text comprehension involves working memory capacity to maintain and process bigger amounts of information in both the L1 (e.g., Schroeder, 2014) and the L2 (e.g., Shin, 2020), and it requires monitoring when coherence is disrupted (Perfetti et al., 2013).

Accordingly, the main aim of the present study is to investigate the high-level cognitive processes of inferential evaluation and revision during L1 and L2 text comprehension. Here we expect, on the one hand, longer gaze duration and/or go-past time in the unexpected ("*grill*") compared to the expected ("*oven*") target word in both the L1 (Pérez et al., 2016) and the L2, but with a smaller effect in the latter (Pérez et al., 2019; Ushiro et al., 2016). This would

suggest a less efficient inferential evaluation in the non-native language. On the other hand, we predict that, after the presentation of the unexpected condition in the text, readers will have longer go-past time and/or total time in the sentence, or longer total time in the target word of the text, when the incongruent concept is presented (“roasted”). This would indicate difficulties to revise the situation model because that concept passively activates the initial prediction (“oven”) causing interference by competition (Kendeou & O’Brien, 2014). Once more, this effect is expected in both L1 and L2 comprehension but with longer times in the latter, signalling less efficient inferential revision in the second language (Pérez et al., 2019; Schleicher & Schwartz, 2022).

In addition, as a second aim, we want to investigate whether individual differences in inhibitory control and linguistic proficiency predict these processes. Whereas we do not expect individual differences of executive control and more specifically of inhibitory control, in the evaluation process during L1 comprehension (Pérez et al., 2015, 2016, 2020), we have no clear predictions regarding the L2, as the literature on evaluation has shown mixed results in relation to both executive control (proactive vs. reactive control) and L2 proficiency (e.g., Pérez & Bajo, 2022; Pérez et al., 2019; Zirnstein et al., 2018). Regarding revision, individual differences in inhibitory control should be associated with revision in the L1 (Pérez et al., 2015, 2016, 2020). Finally, although the literature is a bit uncertain, we also expect individual differences in both executive control and L2 proficiency to be related to revision during L2 comprehension (Pérez et al., 2019).

To test our hypotheses, we used the inferential mismatch detection paradigm developed by Pérez et al. (2016) with eye-tracking to register eye movements (i.e., fixations). Eye movements were preferred for several reasons (see also Hyönä & Kaakinen, 2019): 1) compared to other behavioural measures (e.g., reading times for a whole sentence) it allows us to capture the online processing of evaluation and revision; 2) it reflects the natural way in which comprehension is happening (freely moving across the text), avoiding the frequent artificial word-by-word presentation of EEG studies; and 3) it permits to distinguish between early and late processing (early and late fixations), which provide valuable information about when high-level comprehension processes occur.

2. Method

2.1. Participants

Twenty-two native English speakers and twenty-two native Spanish speakers who were highly proficient in English (their L2) were recruited by an advertisement placed on the website of the University of Cambridge (UK). This means our participants were either studying at or working for this institution, which ensured a

comparable socio-economic status and intellectual capacity. Two L2-English and one L1-English speakers were discarded due to excessive noise in the eye movements’ register, so final analyses were based on twenty-one native English participants (14 females; age: $M = 21.33$, $SE = 0.85$; range 18–34) and twenty Spanish-English participants (13 females; age: $M = 28.10$, $SE = 0.97$; range 20–36). A recent study using the same paradigm and a similar sample size reported sufficient statistical power (Wigdorowitz et al., 2023).

Only L2-English speakers who had lived in an English-speaking country for at least one year ($M = 2.78$, $SE = 0.37$, range 1–7) and considered themselves to have a high level of English proficiency were invited to the study (see Table 2 for more details). The age of L2 acquisition was about 9 years-old ($M = 9.20$, $SE = 0.95$, range 3–20), and many of these non-native participants had learned at least a third language (65% of multilinguals compared to 35% of purely bilinguals). In addition, many of the native participants had acquired at least a second language (62% with at least two languages compared to 38% of purely monolinguals). There were no known reading disabilities and normal or corrected to normal vision.

Informed consent was obtained for all participants prior to testing and a monetary compensation was provided (£15–20) depending on the time of participation. The study was approved by the Ethics Committee of the School of Humanities and Social Sciences of the University of Cambridge.

2.2. Materials

Eye movements were registered to assess high-level comprehension processes. In addition, participants completed a language background questionnaire, a linguistic proficiency task and an inhibitory control task. Scores of the last two were used as indices for individual differences.

High-level comprehension processes task

We used the inferential mismatch detection paradigm (Pérez et al., 2016) to assess the processes of inferential evaluation and revision in English (L1 or L2). The original text sample (30) was increased to improve statistical power. Concretely, each participant was presented with a total of 68 (4 practice, 64 experimental) three-sentence narrative texts (see Table 1; see Appendix 1 for the full set of materials). The first two sentences prime an inference (oven). Subsequently, the third sentence brings one of two target words: a) the expected, in which the concept primed by the previous sentences is presented and thus it is consistent; or b) the unexpected, in which a plausible but improbable concept appears (“grill”) and therefore is inconsistent with the previous inference, demanding inferential evaluation. Eye movements (gaze duration, go-past time and total time) were

Table 1. Example of story presented in the inferential mismatch task.

Story information and conditions	Structure and measure	Cognitive process
It was the 25th of December and Sophie was back home. As a special treat, her father was making her a traditional Christmas dinner. The turkey was cooking, and it needed another hour in the <u>oven/grill</u> before it was done. Expected/Unexpected	Text <i>Eye movements</i>	Inferential evaluation
The turkey needed to be <u>roasted/barbecued</u> for one more hour. Congruent/Incongruent	Sentence <i>Eye movements</i> <i>Accuracy</i>	Revision

recorded on the whole text but analysed on the target word of the text. The two target words were controlled for length in number of characters ($M = 5.94$, $SE = 0.23$ and $M = 6.03$, $SE = 0.24$ for the expected and unexpected word respectively, $t(63) = -0.34$, $p = .74$), SUBTLEX word frequency ($M = 41.46$, $SE = 6.47$ and $M = 38.82$, $SE = 9.18$, $t(63) = 0.28$, $p = .78$) and age of acquisition ($M = 5.06$, $SE = 0.19$ and $M = 5.16$, $SE = 0.19$, $t(63) = -0.37$, $p = .72$). Importantly, several prior norming studies indicated that native English speakers activated the expected (oven) but not the unexpected (grill) concept, after reading the context of the text in their L1 (see Pérez et al., 2015, 2016).

Moreover, a comprehension sentence was presented below the main text, at the same time. Participants were instructed not to read this sentence before they had read the main text. The comprehension sentence contained one of two conditions: a) the congruent, in which the information is associated with the target concept presented in the main text (e.g., “barbecued” after “grill”); or b) the incongruent, in which the information does not match with the presented target concept (e.g., “roasted” after “grill”). Participants were instructed to press “Yes” if they thought the critical sentence was correct, or “No” if they thought it was incorrect. Importantly, the most difficult condition was when coming from the unexpected condition and finding the incongruent sentence (“grill” → “roasted”), as participants were required to discard their previous interpretation (oven), which was subsequently reactivated (by “roasted”), demanding stronger revision of the situation model. Eye movements (go-past time and total time) were analysed in the target region of the comprehension sentence. Once more, the two target concepts of the sentence were controlled for length in number of characters ($M = 14.95$, $SE = 0.83$ and $M = 14.92$, $SE = 0.82$ for the congruent and incongruent concepts respectively, $t(63) = -0.09$, $p = .93$), SUBTLEX word frequency ($M = 133.94$, $SE = 59.15$ and $M = 156.54$, $SE = 54.00$, $t(63) = -0.36$, $p = .72$) and age of acquisition ($M = 5.49$, $SE = 0.22$ and $M = 5.15$, $SE = 0.22$, $t(61) = 1.21$, $p = .23$). An additional norming study was carried out to see whether the target word of the text was related to the target concepts of the sentence. Participants were presented with one of the two target words of the text (e.g., “grill”) followed by the congruent or incongruent information of the sentence (e.g., *to barbecue*). They were instructed to mark from 1 (Not at all) to 5 (Extremely), how well they thought the two concepts were related. Sixteen participants completed this study, four in each list version combining the four crossed conditions. Means (and standard deviations) for each condition were: expected-congruent = 4.24 (0.65); unexpected-congruent = 4.23 (0.81); expected-incongruent = 1.70 (0.66); and unexpected-incongruent = 1.77 (0.70). A repeated measures ANOVA showed a significant main effect of condition, $F(3, 204) = 297.06$, $p < .001$, $\eta_p^2 = .81$, where post-hoc comparisons with Bonferroni corrections demonstrated differences between the two congruent and the two incongruent conditions ($ps < .001$, in all cases), but no significant differences between the two expected and the two unexpected conditions ($ps \approx 1.00$, in all cases). Thus, this norming study confirmed a congruency relationship between the target words of the text and the concepts of the sentence. Finally, accuracy in response to the comprehension sentence was also assessed as an offline behavioural measure, to understand whether participants had fully comprehended the text.

Language background

An adapted version of the Language History Questionnaire (LHQ, Li et al., 2006) was used to measure language background, self-

assessed L2 abilities in reading, writing, speaking and listening, daily language use and frequency of exposure to friends who were English or Spanish speakers (see Table 2). Differences between the L1 and the L2 in Spanish-English participants were non-significant in the percentage of friends who were native English or Spanish speakers, $t(19) = 1.42$, $p = .17$, indicating they had a similar number of native speaker friends in both languages. However, the percentage of daily language use showed significant differences between L1-Spanish and L2-English, $t(19) = 5.59$, $p < .001$, demonstrating that these participants used their L2 more frequently than their L1 throughout the day.

Linguistic proficiency

Proficiency was assessed with a verbal fluency task, which is considered a measure of lexical retrieval efficiency that requires semantic memory and executive control (i.e., working memory, response inhibition and conflict monitoring; see Giovannoli et al., 2023, for a review). Participants were given a category name, and had to name exemplars from this category within 60 seconds. Subsequently, participants heard a tone and the word “STOP” appeared on the screen for 1500ms. Participants performed the task in English (L1 vs. L2). There was one practice category at the beginning of the task (“furniture”) and two experimental categories (i.e., body parts/professions, colours/fruits and vegetables, and animals/clothes) counterbalanced across participants. Verbal fluency scores were calculated as the average of correctly named exemplars in the two categories. This worked as a LINGUISTIC PROFICIENCY INDEX, where higher scores meant higher proficiency in English. As expected, a *t*-test comparison showed significant differences between the two languages, $t(39) = 7.00$, $p < .001$, with higher linguistic proficiency in L1-English ($M = 24.07$, $SE = 0.97$) than in L2-English ($M = 15.35$, $SE = 0.77$)². The linguistic proficiency index was included in the analyses to explore individual differences.

Inhibitory control

We used the flanker task developed by Luk et al. (2010) to measure inhibitory control³. Here, participants are required to press a key on the left side of the keyboard (Z) if the direction of a red target chevron is pointing to the left, and a key on the right side (M) if the target is pointing to the right. The task contained three types of trials: BASELINE TRIALS where the target chevron appears alone; NEUTRAL TRIALS where the target chevron is flanked by black diamonds so not causing interference with the target; and EXPERIMENTAL TRIALS where the target chevron is flanked by four black chevrons. Two possible conditions are presented in

Table 2. Means (and standard errors) for variables related to self-assessment in L2 proficiency and L1 vs. L2 language exposure, in Spanish-English participants.

	L1-Spanish	L2-English
Reading		6.00 (0.15)
Writing		5.70 (0.18)
Speaking		5.65 (0.17)
Listening		5.95 (0.15)
Friends (in %)	39.30 (6.20)	57.00 (6.46)
Daily use (in %)	25.70 (4.21)	70.35 (4.03)

Note. Self-assessment of L2 abilities was based on a 7-point scale with 1 = Very poor, 2 = Poor, 3 = Fair, 4 = Neutral, 5 = Good, 6 = Very good to 7 = Native-like.

experimental trials, CONGRUENT, in which the flanking black chevrons point in the same direction as the target, and INCONGRUENT, in which the flanking chevrons point in the opposite direction causing interference. A total of 400 trials (40 baseline, 120 neutral, and 240 experimental, where 120 were congruent and 120 incongruent trials) were presented. The target direction of each type of trial had the same number of left and right chevrons. In addition, in the experimental trials, the target chevron appearance was equally varied across three different positions. Because we were interested in the cost effect of interference suppression we subtracted the reaction times of the incongruent trials from the congruent trials. This provided an INHIBITORY COST index, where more cost reflected lower inhibitory control and vice versa. A *t*-test comparison demonstrated no differences between the two languages, $t(39) = 0.30$, $p = .84$, L1-English: $M = 62.16$, $SE = 4.67$ (range: 28–110) and L2-English: $M = 60.10$, $SE = 5.10$ (range: 29–122), signalling both groups were equal in inhibitory control. Inhibitory cost index was also included in the statistical analyses to investigate individual differences.

2.3. Procedure

Participants attended one session of about 1 (native English speakers) or 1.30 hours (Spanish–English speakers). In this session, all participants were first assessed on verbal fluency (5–10 mins), flanker (15 mins) and LHQ (10–15 mins) tasks, and then performed the inferential mismatch detection task (35–50 mins including calibration and the actual experiment).

The inferential mismatch detection task was administered in three blocks with two breaks of about 1 min. each, to prevent fatigue. In addition, to prevent excessive noise in eye-tracking data, participants were advised to rest their eyes anytime they needed before each story and to try not to blink during reading. Participants triggered the onset of each trial by fixating a box on the left of the screen. Both the main text and the comprehension sentence appeared at the same time and participants read at their own pace, starting with the text. They were instructed to press the “Yes” key if they thought the comprehension sentence was true, or “No” if they thought it was false. The display disappeared when they pressed one of these two keys, which were counterbalanced across participants. The 64 experimental stories were presented to each participant only once. The assignment of conditions was counterbalanced in four lists across participants, so that each participant saw 16 stories in each crossed condition combination of expectancy and congruency (expected-congruent, expected-incongruent, unexpected-congruent, or unexpected-incongruent). Each list was completed by a similar number of participants and the presentation of trials was randomised. Four practice stories presented at the beginning of the experiment ensured that instructions were understood.

2.4. Apparatus

Behavioural tasks were presented with the E-prime software (Schneider et al., 2002) on a 14” screen. The inferential mismatch detection task was administered in Experiment Builder software on a 19” CRT video monitor (refresh rate = 75 Hz). We used the Times New Roman font, with a size of 20, presented in black colour. The background was white for instructions and grey for all (practice and experimental) trials. Texts started in a height location of 300 and a width of 384, whereas comprehension sentences began at a height of 500 and at the same width

of the texts. Eye movements were monitored using an EyeLink 1000 eye-tracker (SR Research; Mississauga, Canada) with a sampling rate of 1000 Hz. A chinrest and forehead rest were used to minimise head movements and to maintain a constant viewing distance of approximately 60 cm. Viewing was binocular but only the right eye was tracked during the experiment. A nine-point calibration procedure was performed to ensure that tracking accuracy was within 1° of visual angle. Participants were required to fixate a drift correction point at the beginning of each trial to start reading the texts. Recalibration was carried out between trials as needed. The extraction of eye movements was carried out using EyeLink’s own Data Viewer software.

2.5. Dependent measures

Eye-movements

Several eye movements (i.e., fixations) were registered in the inferential mismatch detection task: 1) gaze duration, the total duration of all fixations in a region before leaving it from the left or right (early measure); 2) go-past time, the sum of all fixations from first entering a region from the left to exiting it from the right, including rereading of previous parts of the text (intermediate measure); and 3) total time, the total duration of all fixations in a region, including first and second-past times (late measure). Accordingly, inferential evaluation was assessed by gaze duration and go-past time in the text region, so before participants encountered the comprehension sentence. In contrast, revision was tested by go-past time and total time in the sentence region, as well as by total time in the target word of the text region, assuming the latter captured the rereading of the text once the comprehension sentence had been encountered.

Accuracy

Accuracy (1 = correct; 0 = incorrect) in the comprehension sentence was also analysed. Z-scores were extracted to avoid convergence problems.

2.6. Data analysis

All data were checked to ensure that no participant read the comprehension sentence before reading the text at least once. Text 17 (see Appendix 1) was removed from all the analyses due to a near chance percentage of correct answers in the comprehension sentence in both the L1 (43%) and the L2 (54%). The remaining 63 stories obtained relatively high accuracy in both the L1 ($M = 84%$, $SE = 6.22$, range: 66–95%) and the L2 ($M = 81%$, $SE = 7.34$; range: 69–100%). A minimum cut-off of 80ms was applied for all eye movements’ measures (removed data were 0.31%). In addition, extreme outlier eye movement data per language and expectancy (and congruency, when it applied) was detected by using the outliers boxplot tool of IBM SPSS Statistics (IBM Corp, 2019; Version 26), and replaced by the mean for gaze duration (0.40%; range: 85–1664ms.), go-past time (0.36%; range: 91–2174ms.) and total time (0.52%; range: 100–4258ms.) in the text region, and go-past time (0.16%; range: 83–4639ms) and total time (0.55%; range: 101–9206ms.) in the sentence region.

Eye movements were analysed by linear mixed-effects models (LME) with the *lmer* function of the lme4 R package (Bates et al., 2011), whereas accuracy (binomial measure) was analysed through mixed-effects logistic regression models (MELR) with the *glmer* function of the same lme4 R package. Participants and Items were included as random factors, and Language,

Expectancy, Congruency (when it applied), and the centred values of Proficiency and Inhibitory cost were fixed factors. Separate models were conducted for each dependent measure. The full fixed structure of the models assessing inferential evaluation contained two three-way interactions (language x expectancy x proficiency + language x expectancy x inhibitory cost), whereas the models on revision had two four-way interactions (including congruency).

We first looked for the optimal random structure by keeping the maximal fixed structure. The random factors of Participants and Items were kept in all models, with different random intercepts. The random slopes were Language, Expectancy and Congruency (when it applied), which included all possible combinations. Data-driven model comparison (see Pérez et al., 2016, Appendix 2) was used to extract the optimal random structure. Subsequently, we looked for significant fixed effects using stepwise model comparison from the most complex to the simplest model, and selecting the one with lowest AIC and BIC (for the whole rationale see Pérez et al., 2016). *P* values were provided by the *anova* function of the *lmerTest* R package (Kuznetsova et al., 2015). Effects sizes for LME were informed by partial eta square extracted by the *eta_sq* function of the *sjstats* R package (Lüdtke, 2020), whereas for MELR were reported using odds ratio and the 95% confidence interval (CI), by the *lsmeans* function of the *lsmeans* R package (Lenth et al., 2018; see also Pérez et al., 2020). To test two-way interactions, we ran pairwise comparisons within each factor level combination by using the *testInteractions* function of the *phia* R package (De Rosario-Martínez, 2015), with Bonferroni correction, and means of the fixed effects were calculated with the *interactionMeans* function of the same package.

3. Results

We first examined gaze duration and go-past time in the target words of the main text (*oven* vs. *grill*), addressing whether readers evaluated their inferential comprehension. Second, we analysed go-past time and total time in the target concepts of the sentence region (*roasted* vs. *barbecued*) and total time in the target words

of the text, investigating whether readers had revised their previous interpretation. Finally, we analysed accuracy of the final sentence to investigate participants' final text comprehension. Means and standard error values of the dependent variables in both regions are provided in Table 3. Taking into account the large number of results, we focused on the fixed effects of each LME and MELR (ANOVA and summary details of each model are provided in Appendix 2).

3.1. Inferential evaluation

LME models were run with Participants and Items as random factors, and Language (L1-English vs L2-English), Expectancy (expected vs. unexpected) and one of the two individual differences indices (Proficiency or Inhibitory control) as fixed factors, on the target concepts of the text region ("*oven/grill*") for gaze duration and go-past time eye movements.

3.1.1. Gaze duration in the text region

The model performed in gaze duration manifested the main effects of language, $F(1, 41) = 15.34, p < .001, \eta_p^2 = .27$, with longer durations in the L2 than in the L1 ($M_s = 236$ and 292 ms and $SE_s = 11.02$ and 11.29 , for the L1 and L2, respectively), and expectancy, $F(1, 62) = 5.22, p < .05, \eta_p^2 = .08$, with longer durations in the unexpected than in the expected condition ($M_s = 260$ and 279 ms and $SE_s = 9.69$ and 11.60 , for expected and unexpected, respectively). No other effects reached significance (all $p_s > .05$).

3.1.2. Go-past time in the text region

The model ran for go-past time manifested the same main effects of language, $F(1, 41) = 19.87, p < .001, \eta_p^2 = .28$, with longer go-past time in the L2 than in the L1 ($M_s = 249$ and 329 ms and $SE_s = 13.60$ and 13.95 , for L1 and L2 respectively), and expectancy, $F(1, 61) = 17.18, p < .001, \eta_p^2 = .17$, with longer go-past time in the unexpected than in the expected condition ($M_s = 279$ and 320 ms and $SE_s = 12.21$ and 14.43 , for expected and unexpected respectively). In addition, this time the two-way interaction of language and expectancy was also significant, $F(1, 2344) = 11.04, p < .001, \eta_p^2 = .11$, where post-hoc comparisons

Table 3. Means (and standard errors) of eye movements and accuracy measures obtained in the target word of the text (*oven/grill*) and sentence (*roasted/ barbecued*) of the inferential mismatch task, divided by language, expectancy and congruency.

		L1-English		L2-English		
		Expected	Unexpected	Expected	Unexpected	
Text	Gaze duration	233 (11)	251 (13)	289 (11)	307 (13)	
	Go-past time	250 (11)	271 (18)	308 (12)	374 (21)	
	Total time					
	Congruent	866 (91)	942 (96)	1173 (98)	1326 (105)	
	Incongruent	925 (91)	953 (96)	1513 (98)	1482 (105)	
Sentence	Go-past time	Congruent	440 (40)	440 (40)	627 (49)	639 (49)
		Incongruent	458 (40)	466 (40)	669 (49)	682 (49)
	Total time	Congruent	865 (93)	941 (93)	1180 (94)	1331 (94)
		Incongruent	923 (106)	952 (106)	1511 (108)	1475 (107)
	Accuracy	Congruent	0.927 (0.29)	0.866 (0.22)	0.961 (0.33)	0.871 (0.22)
	Incongruent	0.950 (0.31)	0.697 (0.19)	0.926 (0.29)	0.697 (0.20)	

Note. Gaze duration, go-past time and total time are expressed in milliseconds, whereas accuracy is proportion of correct responses in the comprehension sentence.

demonstrated significant differences between conditions in the L2, $\chi^2(1) = 20.11$, $p < .001$, but not in the L1, $\chi^2(1) = 2.17$, $p = .28$. Concretely, the L2 showed longer go-past time in the unexpected compared to the expected condition (see Table 3). More importantly, the interaction of language and expectancy was qualified by a significant three-way interaction including inhibitory control, $F(1, 2334) = 9.96$, $p < .01$, $\eta_p^2 = .10$. No other effects reached significance (all $ps > .05$).

To follow up on the three-way interaction, we divided the data by language. The remaining two-way interaction of expectancy and inhibitory control was not significant in the L1, $F(1, 1168) = 1.06$, $p = .30$, but it was in the L2, $F(1, 1114) = 10.76$, $p < .01$, where higher inhibitory cost (worse control) was associated with longer go-past times in the unexpected compared to the expected condition, whereas lower inhibitory cost (better control) reduced these differences (see Figure 1).

Overall, these results suggest that all participants were able to generate the predictive inference (oven) and subsequently detect a mismatch in the unexpected information (“grill”), indicating a good ability to perform inferential evaluation (gaze duration and go-past time). However, in the L2 this process was determined by executive control, demonstrating that after encountering unexpected information with a previous interpretation, non-native comprehenders with higher inhibitory control spent less time rereading previous parts of the text in the second language than those with lower inhibitory control.

3.2. Revision

New LME models were performed with Participants and Items as random factors, and Language (L1-English vs L2-English), Expectancy (expected vs. unexpected), Congruency (congruent vs incongruent) and one of the two individual differences indices (Proficiency or Inhibitory control) as fixed factors, either on the target concepts of the sentence (“roasted/barbecued”) for go-past time and total time, or on the target words of the text (“oven/grill”) for total time, taking into account the target concepts of the sentence.

3.2.1. Go-past time in the sentence region

The model performed on go-past time only showed the significant main effects of language, $F(1, 54) = 19.93$, $p < .001$, $\eta_p^2 = .27$, with longer go-past times in the L2 than in the L1 ($Ms = 451$ and $654ms$ and $SEs = 37.11$ and 46.13 , for L1 and L2 respectively), and congruency, $F(1, 2387) = 6.07$, $p < .05$, $\eta_p^2 = .003$, with longer go-past times in the incongruent than in the congruent condition ($Ms = 508$ and $540ms$ and $SEs = 37.87$ in both, for congruent and incongruent respectively). No other effects reached significance (all $ps > .05$). The lack of significant effects regarding expectancy suggests that revision was implemented in a later stage of online comprehension.

3.2.2. Total time in the sentence region

The model ran on total time (late measure) manifested the significant main effects of congruency, $F(1, 53) = 6.37$, $p < .05$, $\eta_p^2 = .11$, with longer total time in the incongruent than in the congruent condition ($Ms = 1075$ and $1208ms$ and $SEs = 79.92$ and 97.03 , for congruent and incongruent respectively), and proficiency, $F(1, 40) = 32.83$, $p < .001$, $\eta_p^2 = .45$, where higher English proficiency reduced total time in general. Moreover, these two factors also interacted, $F(1, 40) = 7.29$, $p < .05$, $\eta_p^2 = .16$, where post-hoc comparisons showed that higher proficiency reduced total time in both the congruent, $\chi^2(1) = 16.76$, $p < .001$, and incongruent, $\chi^2(1) = 36.52$, $p < .001$, conditions, but the effect was more pronounced in the latter one. Finally, there was also a significant interaction between expectancy, congruency and proficiency, $F(1, 2358) = 4.63$, $p < .05$, $\eta_p^2 = .002$. No other effects reached significance (all $ps > .05$).

To follow up on the three-way interaction, we divided the data by expectancy. The remaining two-way interaction of congruency and proficiency was not significant in the unexpected condition, $F(1, 37) = 0.82$, $p = .37$, but it was in the expected condition, $F(1, 39) = 11.29$, $p < .01$, where once more, higher proficiency reduced total time in both the congruent and incongruent conditions, but lower proficiency increased the time especially in the incongruent condition (see Figure 2).

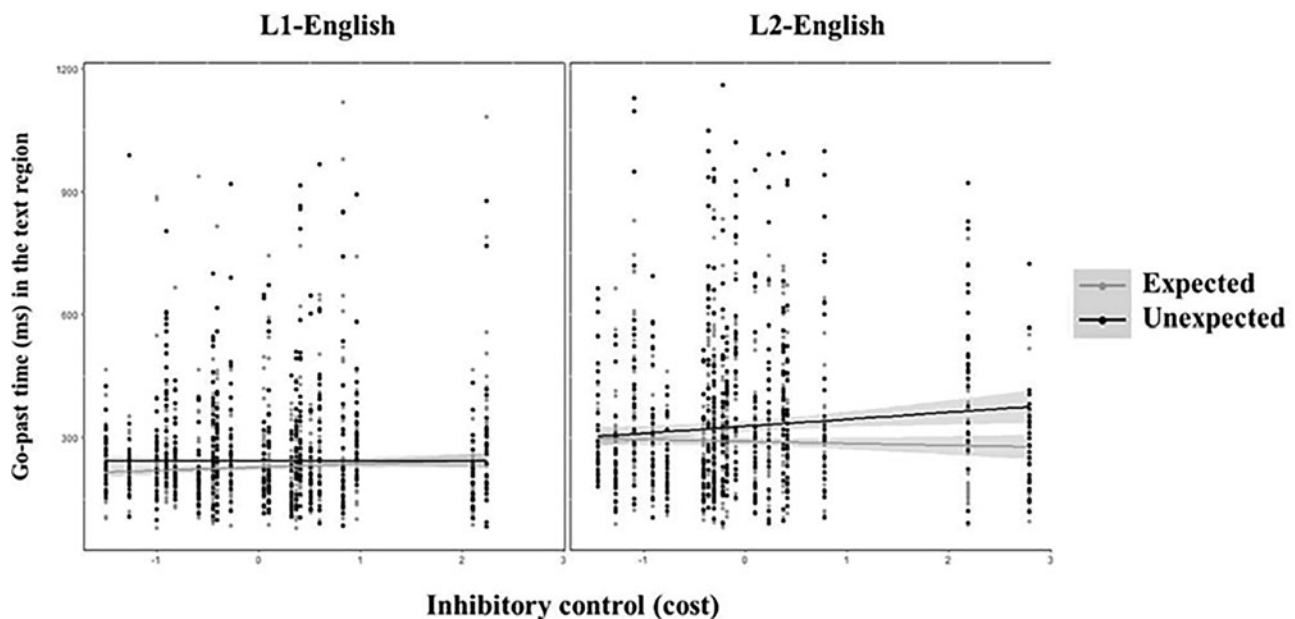


Figure 1. Go-past time (in milliseconds) in the text region, divided by language, expectancy and inhibitory control (cost index).

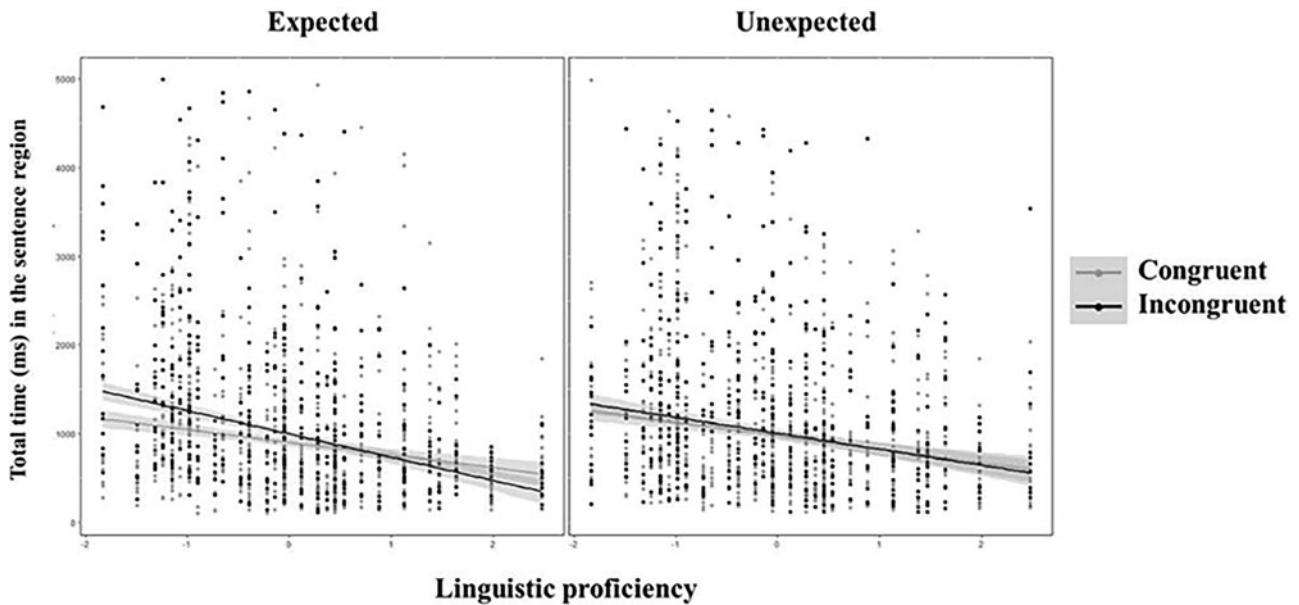


Figure 2. Total time (in milliseconds) in the sentence region, divided by expectancy, congruency and linguistic proficiency.

3.2.3. Total time in the text region

The model performed on the total time spent in the text region demonstrated a complex but very interesting pattern. It showed the significant main effects of language, $F(1, 46) = 17.62$, $p < .001$, $\eta_p^2 = .28$, with longer times in the L2 compared to the L1 ($M_s = 375$ and 554 ms and $SE_s = 33$ and 36 , for the L1 and L2, respectively); expectancy, $F(1, 66) = 68.70$, $p < .001$, $\eta_p^2 = .51$, with longer times after the unexpected condition has been encountered compared to when it was the expected condition ($M_s = 439$ and 633 ms and $SE_s = 29$ and 36 , for expected and unexpected); and congruency, $F(1, 2363) = 10.03$, $p < .01$, $\eta_p^2 = .004$, with longer time in the text region after reading incongruent information in the sentence compared to congruent information ($M_s = 466$ and 521 ms and $SE_s = 30$ and 29 , for both congruent and incongruent). Two two-way interactions were also significant: language and expectancy, $F(1, 2359) = 25.03$, $p < .001$, $\eta_p^2 = .01$, where post-hoc comparisons showed longer times in the L2 compared to the L1 in the unexpected condition, $\chi^2(1) = 32.37$, $p < .001$, but the same effect was only marginal in the expected word, $\chi^2(1) = 4.71$, $p = .06$; and language and congruency, $F(1, 2363) = 5.60$, $p < .05$, $\eta_p^2 = .002$, where longer times after encountering the incongruent compared to the congruent concept in the sentence were found only in the L2, $\chi^2(1) = 14.55$, $p < .001$, but not in the L1, $\chi^2(1) = 0.33$, $p \approx 1.00$. The three-way interactions of language, expectancy and inhibitory control, $F(1, 2361) = 4.67$, $p < .05$, $\eta_p^2 = .002$; as well as expectancy, congruency and inhibitory control, $F(1, 2408) = 5.02$, $p < .05$, $\eta_p^2 = .002$, were significant. More importantly, the four-way interaction between language, expectancy, congruency and inhibitory control was also significant, $F(1, 2382) = 4.96$, $p < .05$, $\eta_p^2 = .002^4$. No other effects reached significance (all $p_s > .05$).

First, to understand whether our sample size was sufficient to defend the four-way interaction, we ran a power analysis using the *simr* package (Green & MacLeod, 2016) with 1000 simulations and alpha set at 0.05. Results showed that the sample size only explained the 62.30% (CI = 59.21–65.31%). Therefore, because this finding is based on insufficient statistical power (at least 80%), any interpretation extracted from this four-way interaction

should be taken with caution. However, we also performed a sensitivity power analysis with $\alpha = 0.05$, $1 - \beta = 0.80$, a numerator df of 1, and a denominator df of 39 in the same four-way interaction, to investigate whether our sample size was sensitive to “the minimum relevant effect size”. The result indicated that the critical F was 4.09, which is indeed smaller than the one that was found here ($F = 4.96$), suggesting our sample size was sensitive enough.

To follow up on the four-way interaction, we first divided the analysis by language. The remaining three-way interaction of expectancy, congruency and inhibitory control was not significant in the L1, $F(1, 1189) = 0.006$, $p = .94$, whereas the same was significant in the L2, $F(1, 1125) = 9.21$, $p < .01$. A power analysis in the latter effect demonstrated sufficient statistical power with our sample size: 88.60% (CI = 86.90–90.50%). Subsequently, for the L2, we divided the analysis by expectancy. The remaining two-way interaction of congruency and inhibitory control was marginally significant in the expected condition, $F(1, 563) = 2.94$, $p = .09$, but it was significant in the unexpected condition, $F(1, 569) = 7.42$, $p < .01$. Post-hoc comparisons in this interaction indicated a relationship between higher inhibitory cost (worse control) and longer times in the text region in the incongruent compared to the congruent condition, $\chi^2(1) = 7.42$, $p < .01$. In contrast, lower inhibitory cost (better control) reduced these differences when the unexpected information was found during L2 comprehension (see Figure 3).

3.3. Text comprehension

A final MELR model was performed with Participants and Items as random factors, and Language (L1-English vs L2-English), Expectancy (expected vs. unexpected), Congruency (congruent vs incongruent) and one of the two individual differences indices (Proficiency or Inhibitory control) as fixed factors, on the accuracy measure to the comprehension sentence.

Accuracy. This model showed the significant main effect of expectancy, $\chi^2(1) = 38.99$, $p < .001$, odds ratio = 4.25 (CI 2.67–6.75), with lower accuracy proportion after encountering the unexpected condition compared to the expected condition

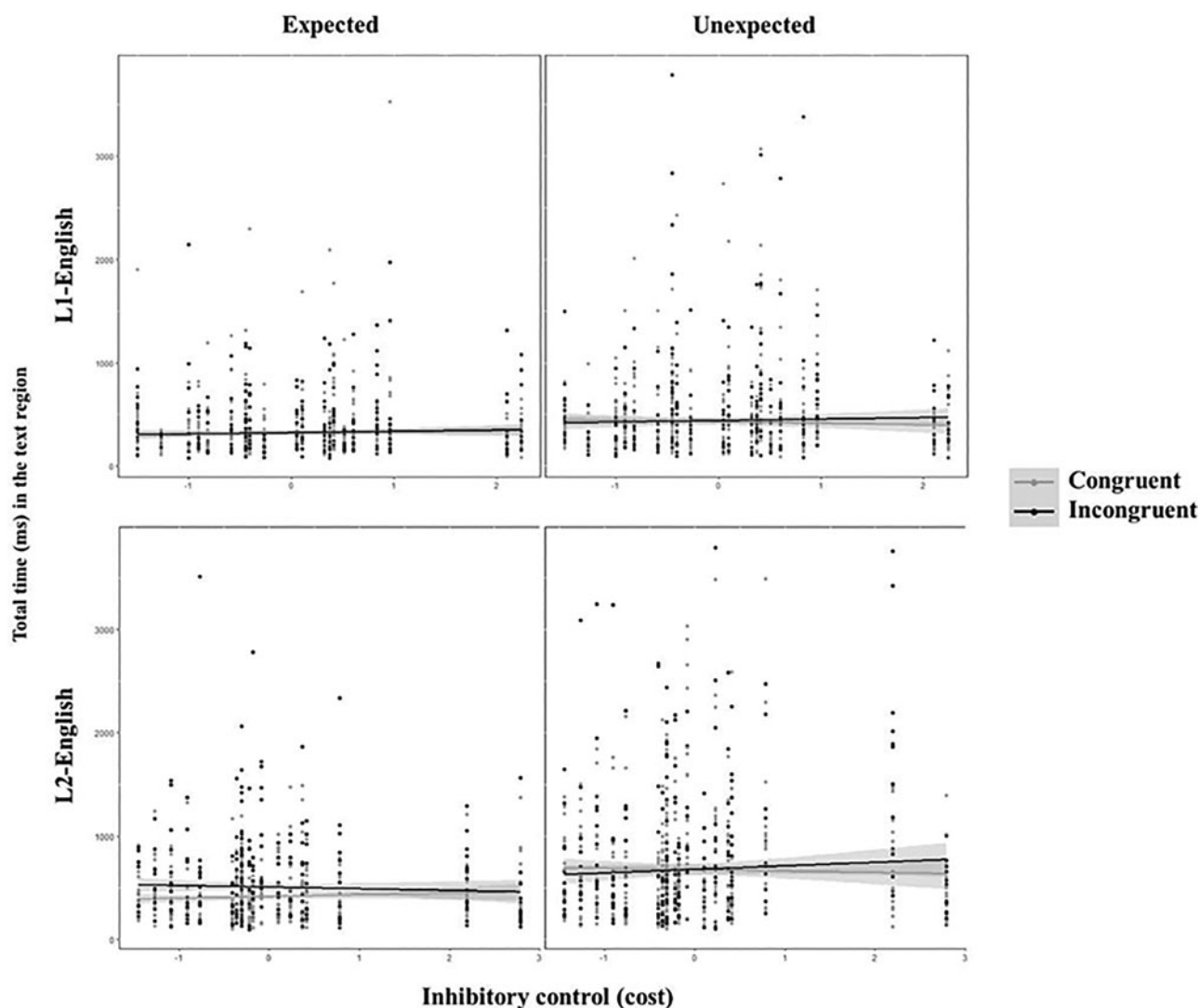


Figure 3. Total time (in milliseconds) in the text region, divided by language, expectancy, congruency, and inhibitory control (cost).

($M_s = 0.94$ and 0.78 and $SE_s = 0.23$ and 0.13 , for expected and unexpected respectively); congruency, $\chi^2(1) = 36.27$, $p < .001$, odds ratio = 1.82 (CI 1.41–2.34), with lower accuracy proportion after reading incongruent information in the sentence compared to congruent information ($M_s = 0.91$ and 0.82 and $SE_s = 0.18$ and 0.17 , for congruent and incongruent respectively); and proficiency, $\chi^2(1) = 9.65$, $p < .01$, where higher English proficiency increased accurate responses in general. The two-way interaction of expectancy and congruency was significant, $\chi^2(1) = 12.40$, $p < .001$, but more importantly, the three-way interaction between language, expectancy and congruency was also significant, $\chi^2(1) = 3.89$, $p < .05$. No other effects reached significance (all $p_s > .05$).

Once more, to follow up on the three-way interaction we divided the analysis by language. This time, the remaining two-way interaction of expectancy and congruency was significant in the L1, $\chi^2(1) = 15.04$, $p < .001$, but not in the L2, $\chi^2(1) = 1.03$, $p = .30$. Specifically, L1 comprehenders showed lower accuracy in the incongruent compared to the congruent sentence only after the unexpected condition was presented (odds ratio = 0.67, CI 0.27–1.67 and 2.80, CI 1.47–5.33, for expected and unexpected,

respectively), which did not occur in L2 readers (odds ratio = 1.97, CI 0.79–4.94 and 2.92, CI 1.52–5.59, for expected and unexpected).

4. Discussion

The main aim of this study was to expand Pérez et al.'s (2016) results on inferential evaluation and revision during L1 comprehension into L2 processing. In addition, inhibitory control and linguistic proficiency were assessed to investigate individual differences. Our findings are discussed in relation to these two aims.

4.1. L1 and L2 inferential evaluation

Using the inferential mismatch detection task, we observed that after the presentation of a context which biased a prediction (see Table 1), there was a time cost effect (longer gaze durations and go-past times) when encountering the unexpected word (“grill”) compared to the expected word (“oven”). This is consistent with most evidence concerning comprehension monitoring (e.g., Cain et al., 2004; Joseph et al., 2021; Kaakinen et al., 2014;

Pérez et al., 2016), where an increase in processing time after the presentation of inconsistent information indicates good comprehension evaluation. This cost effect was also qualified by language, where L2 (but not L1) comprehenders took longer in the unexpected compared to the expected word. However, language differences only occurred in go-past time (the sum of all fixations from first entering a region from the left to exiting it from the right, including rereading of previous parts of the text) but not in gaze duration (the total duration of all fixations in a region before leaving it from the left or right side), suggesting that although both groups detected the inconsistency during early processing, only L2 comprehenders spent longer time (intermediate processing) rereading previous parts of the text. This is in line with studies signalling that although young adults usually detect local inconsistencies in adjacent sentences in the L2, they often fail to do so when the inconsistency is presented globally (e.g., Morishima, 2013; Ushiro et al., 2016, 2021). In our paradigm, inconsistencies were presented almost locally (third sentence); however, they held up on inferencing, and therefore they also involved global coherence. Accordingly, longer rereading times in Spanish–English participants signalled an extra need to monitor inferential comprehension in the non-native language.

Several bilingual studies on predictive processing have shown later onsets and/or weaker effects (i.e., quantitative differences) in the L2 compared to the L1, associated with semantic information (Chun & Kaan, 2019; Dijkgraaf et al., 2019; Kim & Grüter, 2021; Mitsugi, 2020, 2022; Schlenker & Felser, 2021). Our data do not seem to support the idea of a weaker inferential processing in L2 comprehenders, as L2-English participants clearly detected information that was unexpected with respect to the contextual interpretation, demonstrating their ability to perform predictive processing (see Dijkgraaf et al., 2017, for a consistent effect). This finding also contradicts the RAGE hypothesis, by which L2 speakers would have a reduced ability (Grüter et al., 2014, 2017) or a reduced utility (Grüter & Rohde, 2021) to generate predictions. We believe this is due to the type of materials that were used in our study. Specifically, the use of texts (instead of words or sentences) that contain longer information units about the context of the story might allow L2 comprehenders to generate predictions in a native-like manner. However, using very similar stories in a sentence-by-sentence self-paced reading paradigm, Pérez et al. (2019) found that although bilinguals performed predictive processing in the L2 to a certain extent, they did so in a less efficient way. Although the present data cannot refute this hypothesis, it is important to note that our experimental procedure (by means of eye tracking technique) permitted natural comprehension, making possible the rereading of the initial parts of the story even before they encountered the target word, which undoubtedly would have facilitated inferencing. Moreover, L2 evaluation was predicted by individual differences in inhibitory control, where lower inhibitory control was related to longer go-past times in the unexpected compared to the expected word, whereas higher inhibitory control reduced these differences. This effect suggests that L2 comprehenders with better inhibitory control spent less time rereading previous parts of the text after encountering the unexpected information. Some studies have also found a relationship between L2 evaluation and executive control (e.g., Pérez & Bajo, 2022; Zirnstein et al., 2018 – but see Pérez et al., 2019, for null results). For instance, Zirnstein et al. (2018, Exp. 2) observed that advanced Chinese–English bilinguals' ability to detect unexpected information (“*ten*”), in highly predictive sentences (e.g., “*After their meal, they forgot to leave*

a ____ for the waitress”) presented in the L2, was related to inhibitory control. Concretely, bilinguals with faster speed of processing in the AX-CPT task reduced these prediction error costs (smaller late frontal positivity), which they interpreted as a better ability of bilinguals to inhibit the previously formed expectation (“*tip*”). Beyond the methodological and experimental differences, our findings are in accordance with Zirnstein et al.’s (2018) results, demonstrating that L2 comprehenders with better inhibitory control showed less rereading of previous text information after encountering unexpected information, because they implemented more efficient suppression mechanisms when leading with the interference caused by the initial interpretation. In addition, in Zirnstein’s study prediction costs were also associated with linguistic proficiency. In particular, contrary to what the authors expected, prediction costs were not related to verbal fluency in the L2, which was explained by the fact that their participants were highly proficient and immersed in L2-English. However, they did find a relationship between higher L1 verbal fluency and bigger prediction costs (larger frontal positivity) suggesting that better regulatory skills in the L1 helped to generate predictions in the L2. In our study, L2 inferential evaluation was not predicted by individual differences in L2 verbal fluency, confirming Zirnstein et al.’s (2018) findings. Importantly, our L2 participants were also highly advanced Spanish–English bilinguals who were immersed in the L2 at the moment of testing, therefore making our samples fairly comparable. Unfortunately, we did not assess L1-Spanish verbal fluency in the present study, leaving this issue a matter of research for future studies.

Finally, as expected, individual differences either in inhibitory control or linguistic proficiency were not associated with inferential evaluation in the L1. Studies investigating the detection of information that is inconsistent with knowledge-based predictions in young adults suggest executive control (i.e., working memory, cognitive control or inhibitory control) and/or L2 proficiency are not required during L1 text comprehension (Pérez et al., 2015, 2016, 2019, 2020), which is also consistent with the idea that evaluation can be a routine, passive, and nonstrategic process (Kendeou, 2014).

4.2. L1 and L2 revision

The inferential mismatch detection paradigm included a comprehension sentence just below the text (see Table 1). An incongruence effect (longer go-past times and total times) was found after encountering the incongruent (e.g., “*barbecued*” for “*oven*”) compared to the congruent concept (e.g., “*roasted*”), indicating that comprehenders found it more difficult to process sentence information when this was not coherent with the story (Pérez et al., 2016). The same effect was also found in the text (total times in the text region), suggesting that our participants revisited the text more often after encountering incongruent information in the comprehension sentence. Importantly, this effect was modulated by language, demonstrating that L2 (but not L1) comprehenders were more likely to revisit the text. This suggests that L2 comprehenders felt the need to carry out text reanalysis when sentence comprehension was incorrect. In relation to this, some studies have shown that L2 readers may take longer (as seen in response times) to answer incorrect compared to correct questions (e.g., Taguchi, 2005).

Moreover, to understand whether participants had revised their situation model, we looked at the condition in which readers had found the unexpected word in the text (“*grill*”), and then

compared both congruent (“*barbecued*”) and incongruent (“*roasted*”) concepts encountered in the sentence, predicting longer eye movements in the latter one due to a possible reactivation of the initial prediction (oven). In contrast to Pérez et al. (2016), our results did not show a general effect of revision (expectancy and congruency interaction). Instead, revision was predicted by both inhibitory control and language group. Concretely, individual differences in inhibitory control were not associated with revision in the L1, whereas this was the case for L2 revision, where L2 comprehenders with lower inhibitory control spent longer times in the unexpected word of the text region after encountering the incongruent concept in the comprehension sentence. In contrast, L2 comprehenders with higher inhibitory control spent less time, indicating once more, a better ability to suppress interference coming from the no longer plausible prediction. This is related to studies showing that the no longer valid initial interpretation is maintained, causing interference and disrupting comprehension in both the L1 (e.g., Blanc et al., 2011; Rapp & Kendeou, 2009) and the L2 (Fujita & Cummings, 2021; Pérez et al., 2019; Schleicher & Schwartz, 2022). According to the KReC theoretical model, this means that even though the alternative interpretation can be successfully encoded, activated and integrated into the mental representation, the initial prediction is passively reactivated from long-term memory causing competition, and therefore interference. Interestingly, although this interference effect may occur during both L1 and L2 discourse revision, it is stronger when the to-be-updated information is presented in the L2 (Schleicher & Schwartz, 2022).

Importantly, revision in our study was related to individual differences in inhibitory control, which 1) extends the KReC framework by demonstrating that the mechanism underlying interference reduction during the competing activation principle involves inhibition, and 2) provides further evidence on the relationship between L2 revision and executive control (Pérez et al., 2019). Readers need to inhibit the initial interpretation if they want to avoid or reduce semantic interference. In fact, another study assessing L1 comprehension demonstrated that rather than being related to a more general executive control mechanism (working memory), revision is specifically associated with inhibitory control (Pérez et al., 2020). However, even though studies in this area have consistently demonstrated the role of executive control in L1 revision (Pérez et al., 2015, 2016, 2020), this relationship was not statistically significant in our study. We believe this is a matter of quantity rather than quality: both L1 and L2 participants showed a similar pattern in revision (see the unexpected condition in Figure 3), but only those experiencing greater interference (participants reading in the L2) reached significance. Nonetheless, further research is necessary to clarify this issue.

Finally, higher proficiency was generally associated with shorter (total) times reading the sentence. Moreover, this general effect was determined by congruency, demonstrating that higher proficiency reduced the time of the sentence in both the congruent and incongruent concepts, but this effect was especially true for the incongruent condition. In turn, this finding was also modulated by a three-way interaction with expectancy, where the relationship between higher proficiency and shorter times in the incongruent sentence was specifically found after the expected (but not the unexpected) target word had been presented in the text. Although these findings do not seem to show a direct connection between linguistic proficiency and the process of revision (see Pérez et al., 2019, for opposite results), they suggest that higher (L1 or L2, depending on the group) verbal fluency was

related to better comprehension of the story when this was generally coherent (“*oven*”), which subsequently translated into faster reading times in the comprehension sentence, signalling a more efficient way to deal with incorrect questions. Verbal fluency has been linked to both verbal abilities such as lexical knowledge and lexical retrieval speed (e.g., Sauzéon et al., 2011; Shao et al., 2014) as well as executive functions such as updating and inhibition (e.g., Henry & Crawford, 2004; Shao et al., 2014). In our opinion, our data reflect how higher vocabulary size and/or better lexical access speed improved comprehension at the situation model level. In fact, higher verbal fluency was also translated into more accurate responses to the comprehension sentence, confirming the importance of verbal fluency in general text comprehension. Interestingly, some studies have found a relationship between verbal fluency and better predictive processing in reading comprehension (e.g., Federmeier et al., 2010), which specifically suggests that readers with higher verbal fluency were better at generating predictive inferences.

4.3. L1 and L2 text comprehension

In the inferential mismatch task, participants were instructed to press “Yes” if they thought the comprehension sentence was correct, or “No” if they thought it was incorrect (see Table 1). The model on accuracy manifested lower accuracy after the presentation of both the unexpected word in the text region and the incongruent concept in the sentence region, compared to the expected word and the congruent concept, respectively. In addition, these two factors interacted, signaling that lower accuracy for the incongruent compared to the congruent concept was only true once the unexpected (but not the expected) word had been encountered. More importantly, language also qualified this interaction, indicating that lower accuracy after the presentation of the most difficult unexpected-incongruent condition was only found in participants comprehending in the L1, but not in those comprehending in the L2. Because differences were specifically found in the condition requiring stronger revision (“*grill*” → “*roasted*”), it suggests that the longer times that L2 comprehenders spent – rereading the text when dealing with the interference from the initial interpretation (total times in the text region) – translated into better text comprehension (higher accuracy).

Prior effect is an interesting but novel result that could be reflecting a cognitive advantage for being immersed in a non-native language environment. That is, a crucial difference between the L1 and L2 group was that, even though many of our native English speakers were also bilinguals (see Participants section), at the time of the study they were living in a L1 context, whereas the non-native Spanish–English speakers had been living in a L2 environment for at least one year ($M=2.78$). L2-immersion is supposed to constantly involve the suppression of the more proficient native language (here Spanish), a situation that is undoubtedly more challenging than a non-immersed context (Zirnsstein et al., 2018, 2019). Importantly, it has been suggested that the ability to regulate the native language when immersed in a L2 context can bring cognitive benefits in processes such as predictive processing during L2 comprehension (Zirnsstein et al., 2018). Accordingly, our results would indicate that the suppression training, experienced by our non-native comprehenders for being immersed in a L2 environment, was transferred into better suppression of the interference caused by the initial interpretation, and therefore translated into better text comprehension. Nonetheless, we would like to acknowledge that although the

interaction on accuracy was significant, this effect was nearly marginal ($p = .049$). In fact, means were very similar between both groups (see Table 3), so once more, further research is necessary to determine how reliable this effect is.

5. Conclusions

To sum up, the present study extends Pérez et al.'s (2016) results on high-level cognitive processes during L1 text comprehension into L2 processing. Highly proficient L2-English speakers demonstrate less efficient online inferential evaluation and revision than L1-English speakers. Individual differences in inhibitory cost modulate both effects, signaling that, once readers encounter information that is unexpected with their situation model of the story, L2 comprehenders with higher inhibitory control carry out functional repairing processes (such as rereading previous information and revisiting the text) to solve the inconsistency, compared to lower inhibitory control L2 readers. This seems to be based on a better ability to inhibit interference coming from the initial prediction. In contrast, these differences are not observed in L1 comprehenders, suggesting L1 text comprehension is less cognitively effortful. Moreover, higher linguistic proficiency (verbal fluency) predicts better text comprehension in both language groups, especially when the story is fully coherent, suggesting a better ability to generate predictive inferences. Nevertheless, further research is required to understand the complex interaction between L1 and L2 text comprehension, inhibitory control and linguistic proficiency.

Acknowledgements. This research was supported by the grant with ref. 754446-AT3i funded by a MSCA-COFUND Athena3i and the research project with ref. PSI 2017-89324-C2-2-P funded by MCIN, both to Ana Pérez. Funding for open access charge: Universidad de Granada.

Competing interests. The authors declare none.

Open Practices Statement. All data for the study reported here are available in the [www.osf.io](https://osf.io/amdc8/?view_only=7aa261b192e74a098b4206b5b7256e8d) using the following link: https://osf.io/amdc8/?view_only=7aa261b192e74a098b4206b5b7256e8d

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S136672892300069X>

Notes

¹ Notice irony requires inferencing (e.g., Colston & Gibbs, 2002).

² Notice that this difference cannot explain possible effects found exclusively within the L2 group.

³ A monitoring component (based on overall reaction times) has been reported in previous studies using the flanker task (e.g., Costa et al., 2008). Although, the statistical power of the present study does not allow for extra analyses, future studies should try to include this monitoring measure to predict the evaluation process.

⁴ The same four-way interaction was significant in a measure of regressions (looks back) into the target words of the text (specific revision measure), manifesting similar results. With the purpose to simplify the results section we decided to report just total time.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Grothendieck, G. (2011). Package 'lme4'. *Linear mixed-effects models using Eigen and S4 classes. R package version, 1*.

Blanc, N., Stiegler-Balfour, J. J., & O'Brien, E. J. (2011). Does the certainty of information influence the updating process? Evidence from the reading of news articles. *Discourse Processes, 48*, 387–403.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*, 489–503.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology, 96*, 31.

Chun, E., & Kaan, E. (2019). L2 Prediction during complex sentence processing. *Journal of Cultural Cognitive Science, 3*, 203–216.

Colston, H. L., & Gibbs, R. W., Jr. (2002). Are irony and metaphor understood differently? *Metaphor and Symbol, 17*, 57–80.

Cook, A. E., & O'Brien, E. J. (2017). Fundamentals of inferencing during reading. *Language and Linguistics Compass 11*, e12246.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59–86.

De Rosario-Martínez, H. (2015). Phia: Post-hoc Interaction Analysis. R Package Version 0.1–3. Retrieved from <https://cran.r-project.org/>.

de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 373.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition, 20*, 917–930.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience, 34*, 881–900.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*, 149–161.

Fujita, H., & Cunnings, I. (2021). Reanalysis processes in non-native sentence comprehension. *Bilingualism: Language and Cognition, 24*, 628–641.

Giovannoli, J., Martella, D., & Casagrande, M. (2023). Executive functioning during verbal fluency tasks in bilinguals: A systematic review. *International Journal of Language & Communication Disorders*.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*, 493–498.

Grüter, T., & Rohde, H. (2021). Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics, 42*, 51–75.

Grüter, T., Rohde, H., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In *Proceedings of the annual Boston university conference on Language Development*.

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism, 7*, 199–229.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology, 18*, 621.

Hyönä, J., & Kaakinen, J. K. (2019). Eye movements during reading. In *Eye Movement Research* (pp. 239–274). Springer, Cham.

IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1–11.

Joseph, H., Wonnacott, E., & Nation, K. (2021). Online inference making and comprehension monitoring in children during reading: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 74*, 1202–1224.

Kaakinen, J. K., Olkonniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes, 51*, 287–311.

Kaan, E., & Grüter, T. (2021). *Prediction in second language processing and learning* (Vol. 12, p. 1). John Benjamins Publishing Company.

- Kendeou, P. (2014). Validation and comprehension: An integrated overview. *Discourse Processes*, 51, 189–200.
- Kendeou, P., & O'Brien, E. J. (2014). The Knowledge Revision Components (KReC) framework: Processes and mechanisms. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 353–377). The MIT Press.
- Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition*, 43, 133–154.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, 2, 734.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, 1, 3.
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, 38, 202–210.
- Lüdtke, D. (2020). Sjstats: Statistical Functions for Regression Models (Version 0.18.0) <https://CRAN.R-project.org/package=sjstats>.
- Luk, G., Anderson, J. A., Craik, F. I., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: response inhibition versus interference suppression. *Brain and Cognition*, 74, 347–357.
- Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, 69, 574–588.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, 51, 297–384.
- Mitsugi, S. (2020). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching*, 58, 323–349.
- Mitsugi, S. (2022). Polarity adverbs facilitate predictive processing in L2 Japanese. *Second Language Research*, 38, 869–892.
- Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, 50, 577–597.
- Pérez, A., & Bajo, T. (2022). Working memory and high-level text comprehension processes en J. W. Schwieter and Z. Wen (Ed.), *The Cambridge Handbook of Working Memory and Language* (Chapter 21). Cambridge University Press.
- Pérez, A., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive Processing*, 15, 405–413.
- Pérez, A., Cain, K., Castellanos, M. C., & Bajo, T. (2015). Inferential revision in narrative texts: An ERP study. *Memory & Cognition*, 43, 1105–1135.
- Pérez, A., Joseph, H. S., Bajo, T., & Nation, K. (2016). Evaluation and revision of inferential comprehension in narrative texts: an eye movement study. *Language, Cognition and Neuroscience*, 31, 549–566.
- Pérez, A., Hansen, L., & Bajo, T. (2019). The nature of first and second language processing: The role of cognitive control and L2 proficiency during text-level comprehension. *Bilingualism: Language and Cognition*, 22, 930–948.
- Pérez, A., Schmidt, E., Kourtzi, Z., & Tsimpli, I. (2020). Multimodal semantic revision during inferential processing: The role of inhibitory control in text and picture comprehension. *Neuropsychologia*, 138, 107313.
- Perfetti, C. A., Stafura, J. Z., & Adlof, S. M. (2013). Reading comprehension and reading comprehension problems: A word-to-text integration perspective. *Unravelling reading comprehension: Behavioral, Neurobiological, and Genetic components*, 22–32.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144, 1002.
- Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, 35, 2019–2032.
- Rapp, D. N., & Kendeou, P. (2009). Noticing and revising discrepancies as texts unfold. *Discourse Processes*, 46, 1–24.
- Rapp, D. N., & Taylor, H. A. (2004). Interactive dimensions in the construction of mental representations for text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 988.
- Sauzéon, H., Raboutet, C., Rodrigues, J., Langevin, S., Schelstraete, M. A., Feyereisen, P., Hupet, M., & N'Kaoua, B. (2011). Verbal knowledge as a compensation determinant of adult age differences in verbal fluency tasks over time. *Journal of Adult Development*, 18, 144–154.
- Schleicher, K. M., & Schwartz, A. I. (2022). Bilingual Discourse Comprehension: The Role of Language Overlap in Updating the Discourse Representation. *Discourse Processes*, 59, 184–208.
- Schlechter, J. (2022). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, 1–15.
- Schlechter, J., & Felser, C. (2021). Second language prediction ability across linguistic domains: Evidence from German. In E. Kaan and T. Grüter (Eds.), *Bilingual Processing and Acquisition: Vol. 12. Prediction in second language processing and learning*. Amsterdam, Philadelphia: John Benjamins Publishing Company, pp. 48–68.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime: User's Guide. Psychology Software Incorporated.
- Schroeder, P. J. (2014). The effects of age on processing and storage in working memory span tasks and reading comprehension. *Experimental Aging Research*, 40, 308–331.
- Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, 5, 772.
- Shin, J. (2020). A meta-analysis of the relationship between working memory and second language reading comprehension: Does task type matter? *Applied Psycholinguistics*, 41, 873–900.
- St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9, 776–787.
- Taguchi, N. (2005). Comprehending implied meaning in English as a foreign language. *The Modern Language Journal*, 89, 543–562.
- Ushiro, Y., Nahatame, S., Hasegawa, Y., Kimura, Y., Hamada, A., Tanaka, N., Hosoda, M., & Mori, Y. (2016). Maintaining the coherence of situation models in EFL reading: Evidence from eye movements. *JACET Journal*, 60, 37–55.
- Ushiro, Y., Ogiso, T., Hosoda, M., Nahatame, S., Komuro, R., & Nishi, T. (2021). How Japanese EFL Readers Maintain the Local and Global Coherence of Protagonist, Intentionality, and Causality in Narratives: Evidence from Eye Movements. *ARELE: Annual Review of English Language Education in Japan*, 32, 65–80.
- Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2023). High-level listening comprehension in advanced English as a second language: Effects of the first language and inhibitory control. *Bilingualism: Language and Cognition*, 1–15.
- Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, 176, 87–106.
- Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2019). Cognitive control and language ability contribute to online reading comprehension: Implications for older adult bilinguals. *International Journal of Bilingualism*, 23, 971–985.